VECTORIZACIÓN

Planteamiento

Regresión Logística

Sea $h(z) = p(z)$. Entonces, maximizar la función de máxima verosimilitud es equivalente a minimizar la siguiente función de costo:

$$\begin{aligned} C(w) &= -\frac{1}{n}\log(L(w)) = -\frac{1}{n}\sum_{i=1}^{n} y_i \log(h(x_i)) + (1-y_i)\log(1-h(x_i)) \\ &= = -\frac{1}{n}(y^T \log(h(Xw)) + (1-y)^T \log(1-h(Xw))), \end{aligned}$$

donde $X \in \mathbb{R}^{n\times(k+1)}$, $w \in \mathbb{R}^{k+1}$ y $y \in \mathbb{R}^n$. Nótese que esta función de costo es convexa.

Obtención del gradiente de la función costo

$$C(w) = -\frac{1}{n}\sum_{i=1}^{n} y_i \log(h(x_i)) + (1-y_i)\log(1-h(x_i))$$

$$C(w) = -\frac{1}{n}\left(y^T \log(h(Xw)) + (1-y)^T \log(1-h(Xw))\right)$$

$$C'(w) = -\frac{1}{n}\left(\frac{y^T}{h(Xw)}\cdot\frac{\partial h(Xw)}{\partial w} + \frac{(1-y)^T}{1-h(Xw)}\cdot\frac{\partial\, 1-h(Xw)}{\partial w}\right)$$

$$C' = -\frac{1}{n}\left(\frac{y^T}{h(Xw)} - \frac{(1-y)^T}{1-h(Xw)}\right)\left(\frac{\partial h(Xw)}{\partial w}\right) \quad fx$$

$$= -\frac{1}{n}\left(\frac{y^T}{h(Xw)} - \frac{(1-y)^T}{1-h(Xw)}\right)\left(f(x)(1-f(x))X\right)$$

$$= -\frac{1}{n}\left(\frac{y^T}{h(Xw)} - \frac{(1-y^T}{1-h(Xw)}\right)\left(h(Xw)(1-h(Xw))X\right)$$

$$= -\frac{1}{n}\left[\frac{Y^T \cancel{h(X_w)}(1-h(X_w))X}{\cancel{h(X_w)}} - \left(\frac{(1-Y)^T h(X_w)\cancel{(1-h(X_w))}X}{1-\cancel{h(X_w)}}\right)\right.$$

$$= -\frac{1}{n}\left[Y^T(1-h(X_w))X - (1-Y)^T h(X_w)X\right]$$

$$= -\frac{1}{n}\; Y^T X - \cancel{Y^T h(X_w)}X - (h(X_w)X + Y\cancel{h(X_w)})X$$

$$= -\frac{1}{n}\; Y^T X - h(X_w)X$$

$$= -\frac{1}{n}\; X\,(Y^T - h(X_w))$$

$$= \frac{1}{n}\; X\,(h(X_w) - Y^T)$$

ADIÓS?

Demostración de igualdad entre las dos funciones

eq1 = eq2

eq2 = $C(w) = -\frac{1}{n}\left(y^T \log(h(Xw)) + (1-y)^T \log(1-h(Xw))\right)$

eq1 = $C(w) = \underbrace{-\frac{1}{n}}_{\text{constante}} \sum_{i=1}^{n} \underbrace{y_i \log(h(x_i))}_{\text{sumaPonderada1}} + \underbrace{(1-y_i)\log(1-h(x_i))}_{\text{Suma ponderada2}}$

dotProduct == SumaPonderada

$a^T \cdot b = \sum_i a_i b_i$

EdgeCases

$y=0 \rightarrow$ No afecta pq al hacer DotProduct da 0

$y=1 \rightarrow$ No afecta porque $(1-y)$

$X_i = w \cdot X$

$h(X_i) = w \cdot X_i$

$h(X_i) = h(wX)$

$a_i = y_i$, $b_i = \log(h(x_i))$

$a_i = 1-y_i$, $b_i = \log(1-h(x_i))$

$$(y^T)\cdot \log(h(wX)) + (1-y)^T(\log(1-h(xW)))$$

$$-\frac{1}{n}(\text{Suma ponderada1} + \text{suma ponderada2})$$